Research Article | Vol. 12, No. 10 / 1 Oct 2021 / Biomedical Optics Express 6496
Biomedical Optics EXPRESS

Hybrid FPGA-CPU pupil tracker

Bartlomiej Kowalski,[1] Xiaojing Huang,[1,2] Samuel Steven,[1,2] and Alfredo Dubra[1,*]

[1]*Department of Ophthalmology, Stanford University, Palo Alto, CA 94303, USA*
[2]*Institute of Optics, University of Rochester, Rochester, NY 14620, USA*
[*]*adubra@stanford.edu*

Abstract: An off-axis monocular pupil tracker designed for eventual integration in ophthalmoscopes for eye movement stabilization is described and demonstrated. The instrument consists of light-emitting diodes, a camera, a field-programmable gate array (FPGA) and a central processing unit (CPU). The raw camera image undergoes background subtraction, field-flattening, 1-dimensional low-pass filtering, thresholding and robust pupil edge detection on an FPGA pixel stream, followed by least-squares fitting of the pupil edge pixel coordinates to an ellipse in the CPU. Experimental data suggest that the proposed algorithms require raw images with a minimum of ~32 gray levels to achieve sub-pixel pupil center accuracy. Tests with two different cameras operating at 575, 1250 and 5400 frames per second trained on a model pupil achieved 0.5-1.5 μm pupil center estimation precision with 0.6-2.1 ms combined image download, FPGA and CPU processing latency. Pupil tracking data from a fixating human subject show that the tracker operation only requires the adjustment of a single parameter, namely an image intensity threshold. The latency of the proposed pupil tracker is limited by camera download time (latency) and sensitivity (precision).

1. Introduction

The human eye is in constant involuntary movement (rotation), even when fixating on a target [1–3]. When involuntary eye movement is extreme in amplitude and angular speed, often with some periodicity, it is referred to as nystagmus [4]. Peak angular speeds in pathological nystagmus can reach in excess of 400°/s [1,5], introducing substantial blur in flood-illumination ophthalmoscope images and distortion in scanning ophthalmoscope images, making the diagnosing and monitoring of eye disease challenging.

Blur and distortion are particularly problematic in adaptive optics ophthalmoscopy, due to its high magnification and small fields of view [6–9]. Strategies for mitigating image degradation due to eye movement in these instruments include reducing image capture or exposure time [10], averaging multiple registered images [11,12], real-time eye movement compensation, or a combination of these [13,14]. The potential for thermal light damage limits the shortening of exposure times because of the necessary increase in retinal irradiance to achieve acceptable signal-to-noise ratio (SNR). This concern can be overcome through the capture of multiple images of the same retinal location, each with lower irradiance, followed by image registration and averaging. Sharper images can be obtained by discarding the raw images most affected by eye movement [11,12], although this might not be acceptable when measuring retina function [15–17] or blood flow [18]. Thus, eye movement compensation via optical means is a more desirable strategy, as demonstrated using retina tracking in scanning ophthalmoscopes in subjects with physiological (i.e., normal) involuntary eye movement [14,13–20]. These methods, however, require manual identification of an initial retinal template frame with modest distortion due to eye movement [14,20–22], which is not possible in subjects with pathological nystagmus.

Purkinje and pupil tracking are alternatives to retina tracking with lower retinal irradiance that are also robust to ocular media opacities. Purkinje image tracking has been used for

#433766 https://doi.org/10.1364/BOE.433766
Journal © 2021 Received 11 Jun 2021; revised 17 Aug 2021; accepted 11 Sep 2021; published 22 Sep 2021

Biomedical Optics EXPRESS

eye movement stabilization [23,24] using analog electronics achieving a remarkable 20 arcsec root-mean-square (RMS) pupil center estimation precision (inverse of repeatability) and, as of today, an unsurpassed 500 Hz closed-loop correction bandwidth. Subject alignment and operation complexity, however, have prevented the adoption of this technology beyond research settings. Pupil tracking using digital cameras [25–28], offers the potential for simpler operation and lower costs, but improved precision and lower latency are required for eye movement compensation in subjects with nystagmus. Some of the most advanced efforts in this direction include: a 400 frames/s head-mounted pupil tracker using a field-programmable gate array (FPGA) with 2 ms calculation latency and ~20 μm precision [29], a central processing unit (CPU) based 560 frames/s pupil tracker with 4 ms calculation latency with 35 μm precision [30,31], and a commercial device (Eyelink 1000 Plus, SR Research, Ontario, Canada) that captures 2,000 frames/s with 1.4 ms nominal calculation latency and 2 μm precision. Here, we present a pupil tracker built with off-the-shelf components aiming at improving the performance and cost of these devices for eventual eye movement stabilization in subjects with nystagmus. It is important to recognize that retina tracking accuracy through pupil imaging, could be fundamentally limited by the fact that the eyeball is not a rigid body [32,33], and that the crystalline lens wobbles in response to saccades [34,35]. This wobble, however, can be corrected through modeling of the lens as a damped harmonic oscillator, after estimation of the undamped angular frequency and damping ratio of each eye [35].

This manuscript is structured as follows. In section 2, we describe optical setups, two cameras, electronics, and algorithms used to estimate pupil position and orientation. In section 3 we present validation and test experiments using three different hardware configurations to explore the precision-latency compromise, before a summary is presented section 4.

2. Methods

The pupil tracker, depicted in Fig. 1, consists of an optical system with infrared illumination that relays the pupil of the eye onto a complementary metal oxide semiconductor (CMOS) camera connected to an FPGA in a computer with a CPU. Three versions of this optical system were tested using two different cameras to achieve different spatial and temporal sampling.

2.1. Illumination

The eye is illuminated with two 940 nm light-emitting diodes (LEDs; SMBB940DS-1100-02, Marubeni, Tokyo, Japan) to the left and right of the lens closest to the eye. This off-axis illumination produces images in which the pupil of the eye appears dark with two vertically aligned corneal reflections (Purkinje images). As it will become apparent later, this vertical alignment mitigates the number of pupil edge pixel candidates affected by the LED Purkinje images. The LED wavelength was selected to keep photoreceptor stimulation to a minimum, as the photopic spectral luminous efficiency function at 940 nm is lower than 10^{-6} [36]. This choice of wavelength makes the pupil tracker compatible with most ophthalmoscopes and psychophysical experiments.

The use of two LEDs, rather than one, spreads the retinal irradiance across two areas with their centers separated by ~25° of visual angle, providing better light safety than using a single LED. For the purpose of calculating the maximum permissible exposure (MPE) using the American National Standard for the Safe use of Lasers (ANSI Z136.1-2014) [37], we considered the most conservative scenario, that is, we assume that LED light is focused onto a spot on the retina. In practice, the emitting area of each LED is not a point, and the LEDs would only be focused on the retina if the subject is accommodating or is myopic. For this scenario, the MPE for 10-30,000 seconds for a 940 nm continuous wave source focused on the retina is 1.16 mW. The power of the LED pair at the eye was kept below this value at all times.

Biomedical Optics EXPRESS

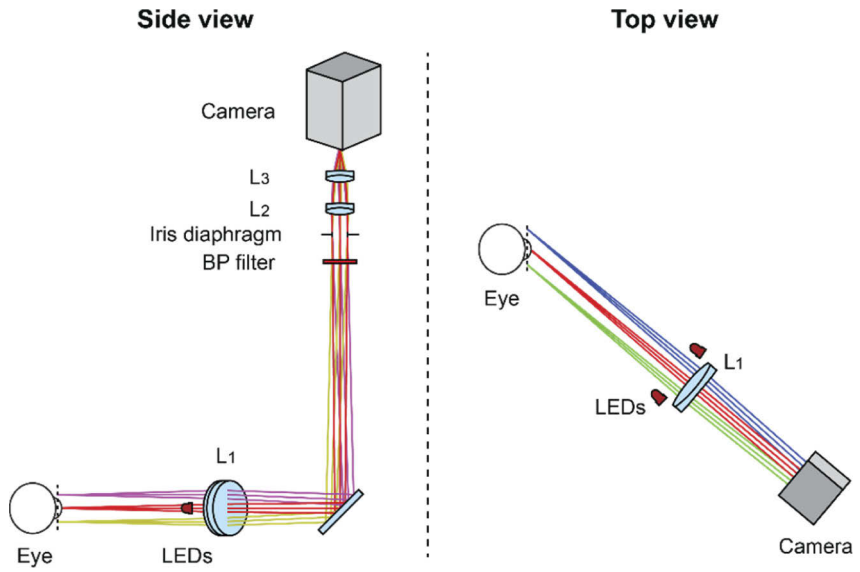

Fig. 1. Pupil tracker optical setup, where BP filter is an interferometric band-pass filter, and L1 to L3 are achromatic doublets. The camera is tilted relative to the optical axis to compensate for the 45° object plane tilt which facilitates integration with ophthalmoscopes or other devices.

2.2. Cameras

The pupil tracker was evaluated with two CMOS cameras with extended-full configuration CameraLink interfaces when capturing 8-bit depth images, to achieve the maximum download data rate that this interface allows. The first camera was an acA2000-340km NIR (Basler AG, Ahrensburg, Germany), with ~15% quantum efficiency at the 940 nm pupil illumination wavelength, 5.5 μm pixel size, and a maximum data bandwidth of 820 Mpix/s, calculated as the product of the camera's internal clock rate (82 MHz) and the 80 bits per clock cycle provided by the 10-tap configuration. The second camera, an Eosens-3CL (Mikrotron GmbH, Unterschleissheim, Germany), with ~5% quantum efficiency at 940 nm, 14 μm pixel size, and a maximum theoretical bandwidth of 710 Mpix/s (Camera Profile 7), when the internal camera pixel clock operates at 75 MHz. The use of the 10-tap CameraLink configuration requires that total number of pixels in the camera region of interest (ROI) is a multiple of 10.

2.3. Optical setups

Three optical setups were used, with an approximate 18 mm square field of view, tilted 45° with respect to the optical axis, and correspondingly tilted image plane (Scheimpflug imaging). For an average human eye, this field of view allows to capture pupils up to 10 mm in diameter [38] with gaze changes of ±20° [39–42]. All three optical setups are telecentric to mitigate pupil size changes due to axial head translation. The "front" of the optical setup consists of the LEDs used for illumination, an achromatic doublet (L_1 in Fig. 1), a fold mirror, an interferometric band-pass filter (84-792, Edmund Optics, Barrington, NJ, USA or FB940-10, Thorlabs, Newton, NJ, USA), and an iris diaphragm that defines the numerical aperture of the system. This front portion of the setup was fixed, allowing quick switching of the "rear" section, which includes two achromatic lenses (L_2 and L_3) and a custom adaptor in which the camera is inserted with a tilt matching that of the image plane. The off-the-shelf lenses and their separation were chosen to achieve the desired magnification, while keeping the imaging performance close to diffraction-limited.

Biomedical Optics EXPRESS

The three optics-camera combinations are referred hereon as: Basler high resolution (-0.18 magnification), Basler high frame rate (-0.085 magnification) and Mikrotron high frame rate (-0.13 magnification). In the high resolution Basler configuration, the camera captured 460×660 pixel images with a 1.7 ms exposure at 575 frames/s, while in the high frame rate Basler setup, 210×300 pixel images were captured with 0.7 ms exposure at 1250 frames/s. The high frame rate Mikrotron setup was used to capture 210×284 pixel images with 0.18 ms exposure at 5400 frames/s.

The optical components (excluding the bandpass filter for simplicity) and their separation along the optical axis for all three configurations are listed in Table 1 and Table 2 below. Image distortion due to the optical setups (i.e., ignoring refraction at the ocular surfaces) across the field of view is ~0.1%. The 100 mm eye clearance was selected to facilitate integration with existing and new ophthalmoscopes, without the need for dichroic mirrors or beam splitters.

Table 1. Separation and tilt of elements in Basler camera optical setups.

		High frame rate			High resolution		
Element	Manufacturer	Part #	Thickness (mm)	Tilt (°)	Part #	Thickness (mm)	Tilt (°)
Eye			100.0	45.0		100.0	45.0
Lens 1	Edmund Optics	49-377	197.4	0.0	49-377	197.4	0.0
Iris	Thorlabs	SM1D12C	9.2	0.0	SM1D12C	12.5	0.0
Lens 2	Thorlabs	AC127-050-B	9.5	0.0	AC254-050-B	26.0	0.0
Lens 3	Thorlabs	AC127-019-B	9.7	0.0	AC254-050-B	13.5	0.0
Camera	Basler	acA2000-340kmNIR	-	5.1	acA2000-340kmNIR	-	10.3

Table 2. Separation and tilt of elements in Mikrotron camera optical setup.

		High frame rate		
Element	Manufacturer	Part #	Thickness (mm)	Tilt (°)
Eye			100.0	45.0
Lens 1	Edmund Optics	49-377	197.4	0.0
Iris	Thorlabs	SM1D12C	13.4	0.0
Lens 2	Thorlabs	AC254-075-B	16.6	0.0
Lens 3	Thorlabs	AC254-030-B	18.2	0.0
Camera	Mikrotron	EoSens-3CL	-	8.5

2.4. Computing hardware

The raw pixel values of the camera are downloaded to a reconfigurable frame grabber (PCIe-1477; National instruments, Austin, TX, USA). This device has a Kintex-7 325T FPGA, that was custom programmed using the LabVIEW FPGA module (National Instruments) and the Vivado Design Suite (Xilinx, San Jose, CA, USA). The frame grabber was installed in a PCIe slot of a computer with an i7-6850K CPU (Intel Corporation, Santa Clara, CA, USA) and a GeForce GTX 1050 discrete graphics processing unit (GPU; Nvidia, Santa Clara, CA, USA). The data flow across hardware components and the algorithm architecture are summarized in Fig. 2, with the FPGA and CPU used for pupil tracking and the GPU for displaying images and data using a custom Python wrapper of the Open Graphics Library (OpenGL, Khronos Group, Beaverton, OR, USA) [43].

Biomedical Optics EXPRESS

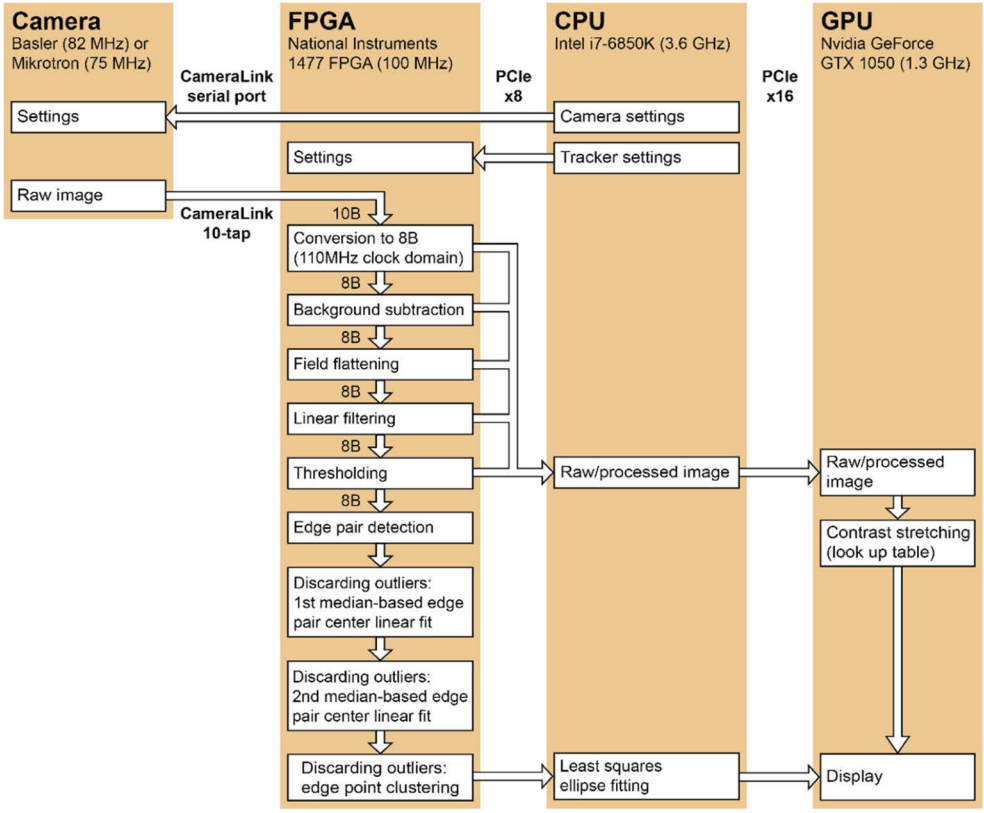

Fig. 2. Hardware and data flow diagram used in the proposed pupil tracker. The converging arrows onto the Raw/processed image indicate that the pupil image can be transferred to the CPU RAM after any of the corresponding processing steps.

The key to achieving pupil tracking with low latency is to process pixel values as soon as they arrive to the FPGA, through what it is often called a "pixel stream." This is faster than the paradigm in which the image processing does not start until the camera image is fully downloaded to a CPU or a GPU. In what follows, the FPGA image processing steps are sequentially ordered so that each new set of pixels arriving to the FPGA with a "tick" of the camera clock, which pushes the previous pixels along the pixel stream. The use of an FPGA allows deterministic timing, because FPGAs only execute programmed processing commands, unlike most CPUs which need to interleave tasks required by the operating system.

2.5. Raw data re-packaging

The first FPGA operation is the re-arranging of the raw pixel values from 80-bit chunks, provided by the CameraLink interface, to the 64-bit chunks used in the pixel stream processing. This re-packaging facilitates the data upload to and retrieval from the FPGA DRAM, needed for the background subtraction and field-flattening described later, as well as transferring image data to the CPU RAM through the PCI-e interface for eventual display. Both the FPGA and camera clocks operate at their maximum respective frequencies, to minimize processing time.

The repackaging was implemented using the LabVIEW component IMAQ FPGA Camera Link Pixel Packer U8 × 10.vi (National Instruments), which must operate at least 1.25 times faster than the camera pixel clock, that is, at 102.5 MHz. Because this is faster than the 100 MHz

Biomedical Optics EXPRESS

FPGA clock, we used a custom FPGA 110 MHz clock domain, a first-in first-out (FIFO) queue in the 110 MHz domain and a FIFO in the 100 MHz domain.

2.6. Background subtraction and field flattening

A background image is generated as the median of a sequence of images collected with the camera ROI, gain and exposure settings to be used for pupil tracking, and with the first lens of the optical setup covered. This background image is stored in the FPGA DRAM and subtracted from every image in the pixel stream after pixel data re-packaging, to mitigate fixed pattern noise typical of CMOS cameras. After this subtraction, negative values are forced to zero.

The background subtraction is followed by a field flattening step which consists of pixel multiplication by a matrix with values equal or larger than one, aiming to compensate for non-uniform illumination, non-uniform pixel sensitivity and vignetting in the optical setup. This matrix is generated by first capturing and averaging a sequence of images of a white flat tilted (45°) piece of paper illuminated with the NIR LEDs. After subtracting the previously estimated background, the resulting image is low-pass filtered using a two dimensional Gaussian filter of user-selectable sigma (width; set to 5 pixels in this work) and normalized by the maximum pixel value. The field flattening matrix is then formed by the inverse pixel values converted to a 16-bit fixed positive decimal numbers in the 0 to 31.999512 range. This data type is a compromise between accuracy, dynamic range, and FPGA resource utilization.

A binary mask, derived from the illumination profile used for field flattening is created to tag the pixels deemed to be too poorly illuminated to be useful for pupil tracking, including those with zero pixel value. This vignetting mask is used by the FPGA logic to exclude these pixels in the pupil edge search.

The background image, the field-flattening matrix and the vignetting mask are all calculated in the CPU, interleaved in 8-byte, 16-byte, and 8-byte chunks, respectively, before uploading to the FPGA's dynamic RAM (DRAM). From there, 64-byte chunks are transferred to and from the FPGA block RAM buffer from which they will be read one 32-byte chunk for every pixel stream clock cycle. In order to cope with the non-deterministic FPGA DRAM access, 32 of these 32-byte chunks are buffered in FPGA memory. After background subtraction and multiplication by the field flattening matrix, the resulting pixel values are output as unsigned 8-bit integers, clamping values larger than 255 to 255.

2.7. Low-pass filtering

Random noise that varies from frame to frame is evident in raw images from both cameras. This is illustrated by the cross-sections of a defocused background-subtracted field-flattened image of a piece of paper in Fig. 3 below (blue curves). In the absence of noise, such line would be horizontal (i.e., zero root-mean-square; RMS). In order to mitigate this noise, the camera images are convolved with a 1-dimensional (1D) Gaussian finite impulse response filter with 15 elements and unit energy after field flattening. This 1D low-pass filtering was selected over 2-dimensional filtering to reduce FPGA code complexity and resource utilization. Median filtering was initially considered, but discarded because its previously thought edge-preserving property has recently been proven incorrect, other than for very high signal-to-noise ratio scenarios [44]. The 1D filter is Gaussian with standard deviation of 4 pixels, which in the images shown in Fig. 3, reduce the image RMS by a factor of 3 (red curves).

It is important to note that the 1D filtering blurs image features, including the pupil edge that we seek to identify for tracking. A benefit of this blur, however, is the smoothening of undesirable edges that might confuse the pupil tracking, such as eyelashes and eyebrows.

Biomedical Optics EXPRESS

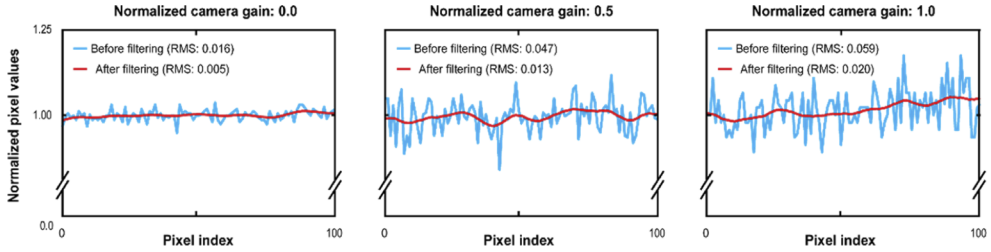

Fig. 3. Cross section of a defocused Basler camera image of a white piece of paper illustrating the amplitude of random noise at three normalized gain values before and after the proposed one-dimensional low-pass filtering.

2.8. *Thresholding*

After the 1D filtering, pixels darker than a threshold intensity are set to one while pixels brighter than the threshold are set to zero, as a first step to identify the dark pixels that are likely to be part of the pupil. This pixel classification relies on the simple observation that when under near-infrared off-axis illumination, the pupil of the eye is darker than the skin, sclera, iris, eyebrows and natural eye lashes in human subjects of all ethnicities. The grey level histograms of images from the front of the eye typically have two broad peaks, with one corresponding to the darker pixels within the pupil, as can be seen in Fig. 4. Here, we adopt the widely utilized strategy of selecting the threshold value as the grey level that corresponds to the minimum between these two broad histogram peaks [30,45,46]. Thresholding, however, does not adequately exclude eye pixels that are dark due to eyelash makeup, which as the lower panels in Fig. 4 show, can be as dark as, or even darker than the pupil pixels.

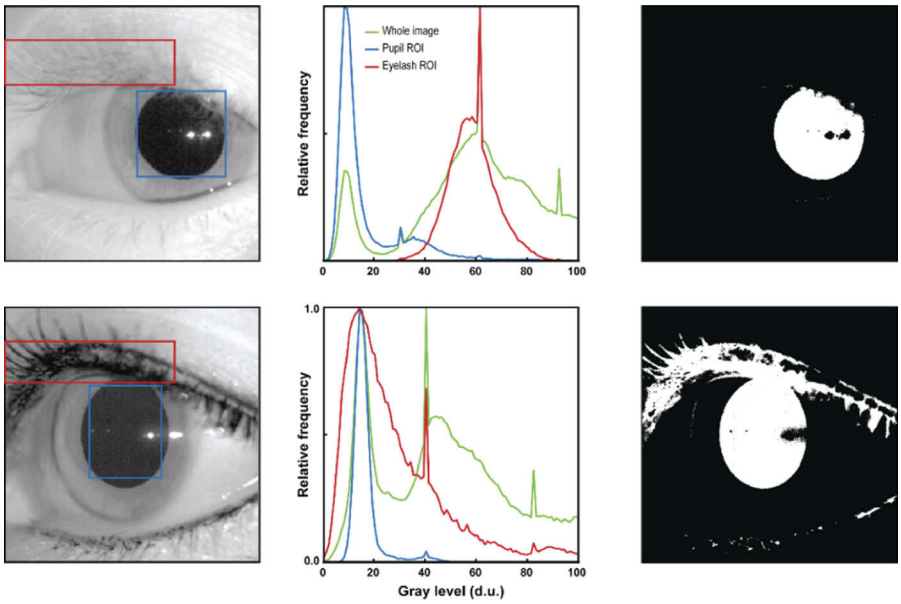

Fig. 4. Images (left), pixel histograms (middle) and thresholded images (right) of the same eye without (top) and with (bottom) eyelash makeup. The makeup shifts the histogram in areas corresponding to eyelashes to darker values, resulting in the thresholding of non-pupil pixels.

Biomedical Optics EXPRESS

2.9. Detection of left-right edge pairs

Following the thresholding, the FPGA pixel stream looks for left-right edge pairs within each line of the binary image, starting from the left side. In this algorithm, a left edge is a one-valued pixel preceded by (the user-defined) n_{out} zero-valued pixels, while a right edge is defined as a one-valued pixel preceded by a left edge and minimum n_{in} consecutive one-valued pixels followed by n_{out} zero-valued pixels. These edges pairs are shown in the sample images of Fig. 5 as red dots. A few pixels at the start and end of each line are excluded from the edge search region. The number of these excluded pixels is the maximum of the radius of the 1D filter and n_{out} columns.

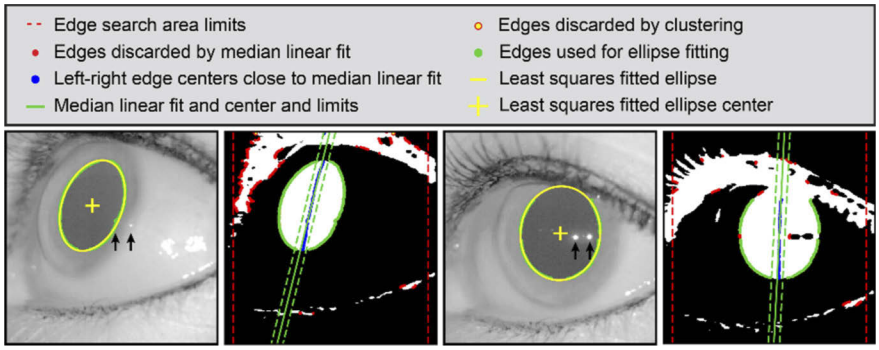

Fig. 5. Binary and raw pupil images (logarithmic intensity grey scale) from a subject wearing dark eyelash makeup that results in the thresholding of numerous non-pupil pixels. The annotations illustrate how the left-right edges are identified for ellipse fitting (green), and which ones are discarded by the median linear fit (red) or the clustering (yellow with red outline, none in this figure). The black arrows indicate the LED Purkinje images (corneal reflections), which in the left images is (incorrectly) classified as a pupil edge because it is close to it.

As the binary images in both Fig. 4 and Fig. 5 show, dark image features such as eyelashes with makeup, can appear in the thresholded binary image, potentially resulting in undesired left-right edge pairs. Two additional algorithms, applied in sequence, seek to remove these undesirable edge pairs. First, we exploit a geometrical property of ellipses that says that the centers of all left-right edge pairs must lie along a line. A robust fit of the edge pair centers by low-pass filtering (7 value-wide average window) their horizontal coordinates is followed by the calculation of the median slope of all consecutive edge pairs (left to right and top to bottom), and the corresponding median intercepts. Then, the edge pairs with centers further than a user-defined distance from the line defined by the median slope and intercept are discarded in a first iteration, before a second iteration with tighter tolerance (see dashed and continuous green lines in Fig. 5). The second algorithm, creates clusters of left and right edges based on their separation, retaining only those with a user-defined minimum number of edge pixels.

2.10. Ellipse fitting

In pupil tracking, the coordinates of the pupil edge pixels are often fit to a circle [46–48] or an ellipse [49,50]. These are mathematically convenient geometrical models that do not account for the irregular inner edge of a pupil but seem adequate for tracking eye movement. In our pupil tracker we fit the coordinates of the pixels identified as pupil edges to an ellipse and use the ellipse center and orientation to track eye movement, defined as the rotation of the visual axis, commonly defined as passing through the pupil center [51]. We assume that the orientation of the ellipse is due to eye rotation around the pupil center (cyclotorsion).

Biomedical Optics EXPRESS

In our model, the column (x_i) and row (y_i) coordinates of each of the N pupil edge points identified in the previous steps are used to fit an ellipse of the form,

$$Ax_i^2 + Bx_iy_i + Cy_i^2 + Dx_i + Ey_i + F = 0. \tag{1}$$

By re-arranging the coordinates of each point as a set of linear equation in the unknowns B/A, C/A, D/A, E/A, and F/A, we get the linear matrix equation,

$$\begin{pmatrix} x_1y_1 & y_1^2 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_Ny_N & y_N^2 & x_N & y_N & 1 \end{pmatrix} \begin{pmatrix} B/A \\ C/A \\ D/A \\ E/A \\ F/A \end{pmatrix} = - \begin{pmatrix} x_1^2 \\ \vdots \\ x_N^2 \end{pmatrix}. \tag{2}$$

We find a least-squares solution to this system of equations by multiplying both sides by the transpose of the first matrix on the left, and then invoking a solve routine that uses the software library for numerical linear algebra LAPACK [52]. The solution is then used to calculate the ellipse semi-major axis (a), semi-minor axis (b), orientation (θ), and center location (x_o, y_o) using the canonical equations [53],

$$\begin{aligned} x_o &= \frac{(BE - 2CD)}{(4CA - B^2)} \\ y_o &= \frac{(BD - 2AE)}{(4CA - B^2)} \\ \theta &= \frac{\pi}{2} + \frac{1}{2}\arctan\left(\frac{B}{A},\ 1 - \frac{C}{A}\right) \\ d_1 &= \sqrt{2\left(x_o^2 + \frac{B}{A}x_oy_o + \frac{C}{A}y_o^2 - \frac{F}{A}\right) / \left(1 + \frac{C}{A} - \frac{B}{A\sin(2\theta)}\right)} \\ d_2 &= \sqrt{22\left(x_o^2 + \frac{B}{A}x_oy_o + \frac{C}{A}y_o^2 - \frac{F}{A}\right) / \left(1 + \frac{C}{A} + \frac{B}{A\sin(2\theta)}\right)} \end{aligned} \tag{3}$$

2.11. Subjects

The study protocol adhered to the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board of Stanford University. Three volunteers with no known ocular pathology were enrolled. The subjects were positioned in front of the pupil tracker using a bite bar attached to a 3-axis translation stage to align and stabilize their head during data collection.

3. Tests and results

3.1. Image dynamic range

Camera gain, camera exposure and/or LED illumination power in pupil tracking are often adjusted to use most of the camera dynamic range. Here, we explore how pupil center estimation is affected when using a reduced dynamic range, for example by lowering the camera gain to achieve lower readout noise or lowering LED power for improved light safety.

As a first experiment, we compare the center of a fitted ellipse in a pupil image of a human subject as we right bit-shift the image (i.e., we divide the pixel values by two, only retaining the integer portion) and the image threshold (initially set to 32). If we consider the center of the

Biomedical Optics EXPRESS

ellipse in the original image as the ground truth, the panels in Fig. 6 show that up to a bit-shift of 4 the ellipse center estimation remains within a fifth of a pixel. If this pixel shift is considered acceptable, then the pupil tracker could be used with a combination of camera gain, camera exposure and LED power that creates images with a dynamic range of just 32 grey levels. When the pupil image spans only 16 gray levels, the ellipse center shifts both vertically and horizontally by more than one pixel. This testing is far from exhaustive, and such a test should be repeated for each experimental condition and choice of algorithm parameters, but it suggests that it is not necessary to capture images with a 256 grey level dynamic range.

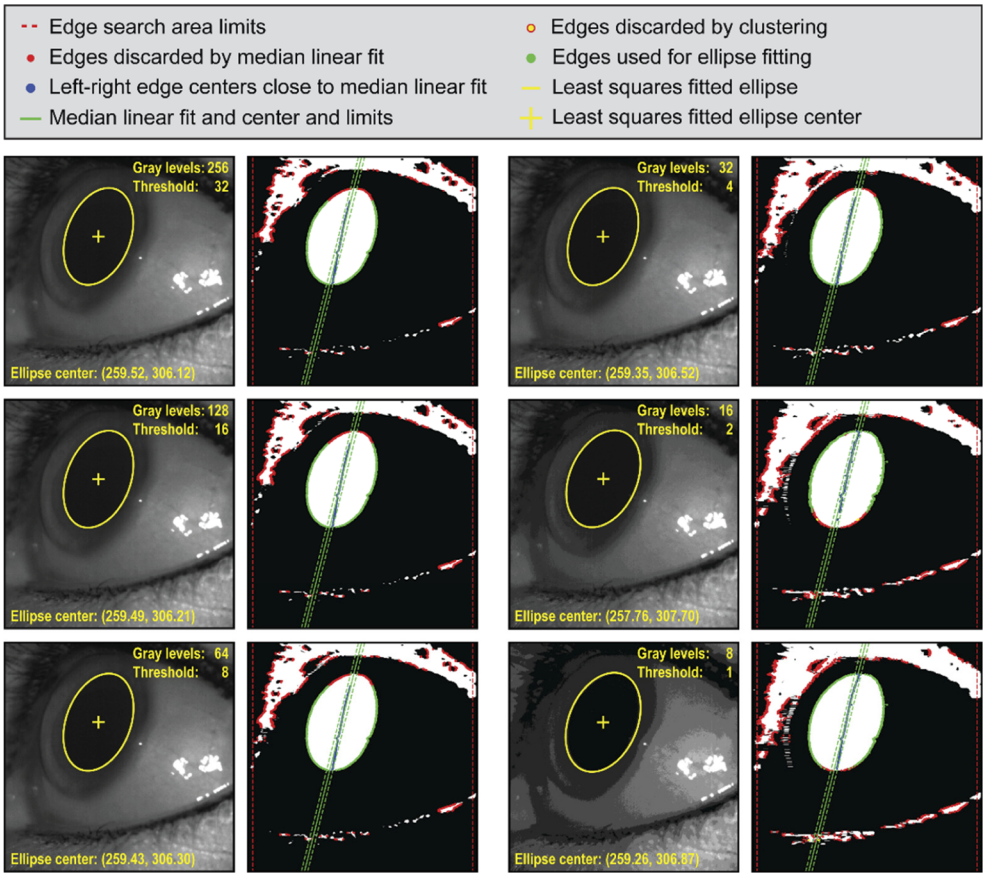

Fig. 6. Raw (linear intensity scale) and binary pupil images from a subject wearing dark eyelash makeup that results in the thresholding of numerous non-pupil pixels. All raw images are generated by bit-shifting (right) the original raw image (top left). The annotations show the edges identified for ellipse fitting (green), and those discarded by the median linear fit (red) and the clustering (yellow with red outline, none in this figure).

In a second experiment, we captured images of an eye changing the camera gain while keeping camera exposure and LED power constant. The resulting raw images, contrast-stretched for display purposes, and the corresponding binarized images after thresholding images are shown in Fig. 7. Ignoring the corneal reflections, the images can be thought of having an approximate dynamic range between 3 and 8 bits, or between 8 and 256 grey levels, respectively. When the image spans only 8 gray levels, the best threshold value to segment the pupil of the eye is 1, which is also the minimum possible. As the corresponding binary image shows, many non-pupil pixels and edges appear, which indicates that the image dynamic range is too small

Biomedical Optics EXPRESS

to separate the pupil from other dark image features, even though most non-pupil edge pixels (red) are correctly discarded. Whenever the raw image spans 32 or more gray levels, there is no noticeable difference in the binarization. As before, this is a very crude test, but it suggests that combinations of camera gain, exposure and illumination power that achieve a minimum image dynamic range of 32, ignoring Purkinje images, are desirable.

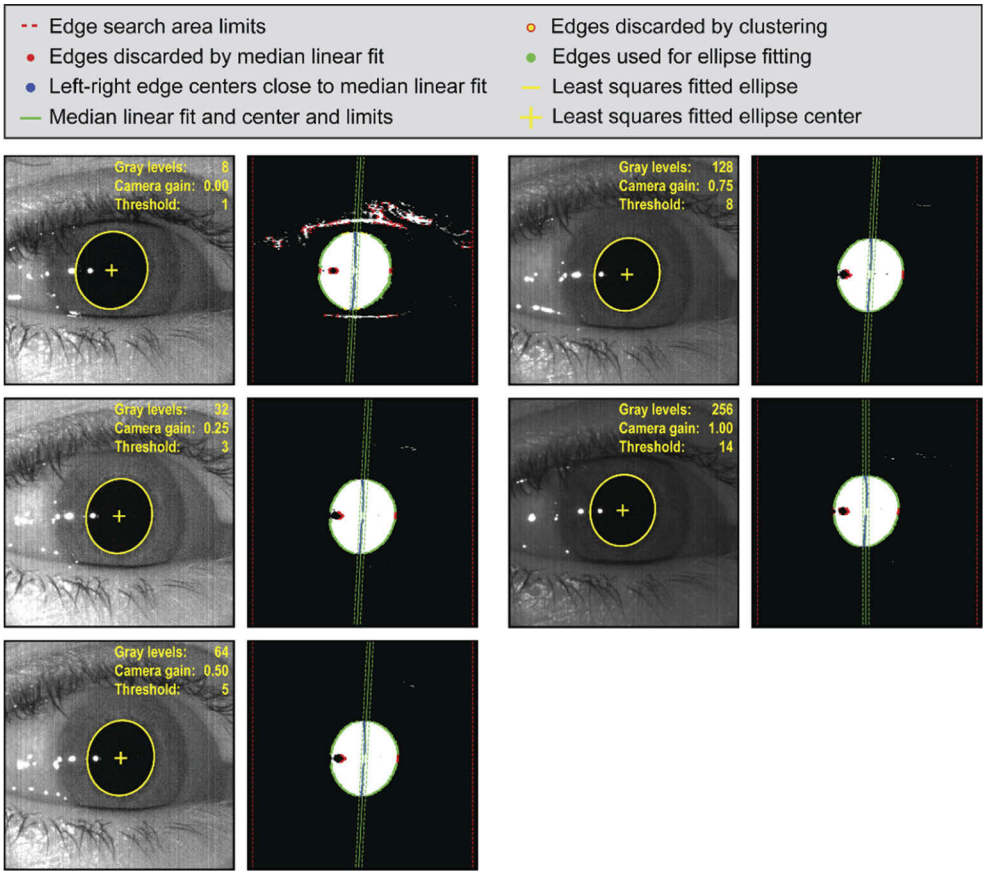

Fig. 7. Raw pupil images (linear intensity scale) captured on the same subject using different camera gains and the corresponding binary images after thresholding. The annotations show the edges identified for ellipse fitting (green), and those discarded by the median linear fit (red) and the clustering (yellow).

3.2. Precision

The precision of the pupil tracker was evaluated by capturing sequences of 100 images of a 6 mm black circle printed on white paper, as a model pupil. These image sequences were captured using all three optical setups using light levels that would make the image span either 64 or 256 gray levels. For the Basler camera, image sequences were captured with normalized gain values 0.0, 0.5 and 1.0. The Mikrotron camera does not allow gain changes. The ellipse parameter repeatability, defined here as the standard deviation across the 100 repeated measurements, is reported in Fig. 8 for all experimental conditions. The conversion factor (0.3438) between microns of pupil movement and arc minutes of estimated rotation was calculated assuming that in an average eye, the pupil is ~10 mm from the center or rotation of the eye.

Biomedical Optics EXPRESS

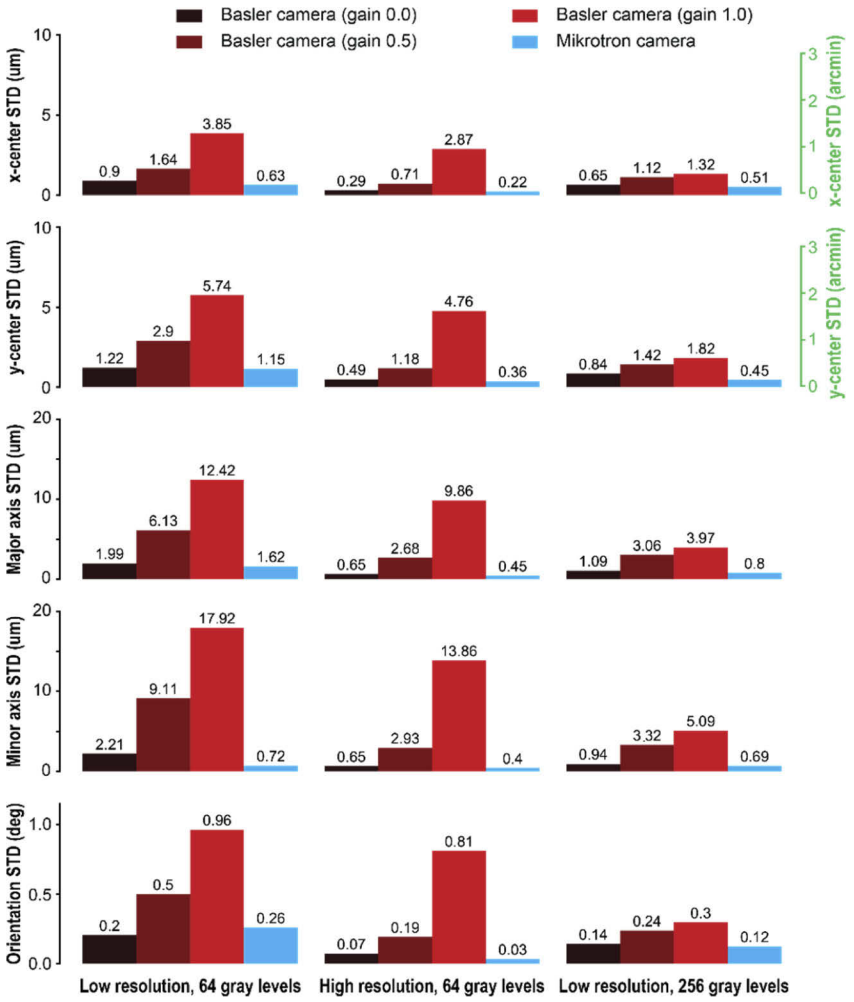

Fig. 8. Repeatability (standard deviation) of fitted ellipse parameters across 100 images of a 6 mm diameter circular model pupil, spanning either 64 or 256 gray levels.

For the Basler camera, it can be seen that the precision worsens with increasing gain, as the amplification of the readout electronics results in higher signal but lower SNR. Also, increasing pupil sampling improves precision, but at the cost of exposing the eye to more light. The overall ellipse parameter precision from the Mikrotron camera images is superior to that of the Basler camera, although due to their difference in quantum efficiency at 940 nm, the Mikrotron camera requires approximately three times higher intensity. In agreement in with the experiments from the previous section, the bar plots also show that 256 gray level images provide superior precision with respect to 64 gray levels, by between a few percent and up to a factor of two.

3.3. Latency

Here we define pupil tracking latency as the time required to: readout the camera sensor, download the pixel values to the FPGA, complete the FPGA calculations which overlap with the image download, transfer the pupil edge coordinates from the FPGA to the CPU, and the ellipse fitting. This definition is independent of the camera exposure, which we choose to maximize frame rate while also minimizing peak optical power delivered to the eye. Timing and latency measurements

Biomedical Optics EXPRESS

summarized in Table 3 below, were performed using an oscilloscope MSO 2024B (Tektronix, Beaverton, OR, USA) using the end of the camera exposure signal as the time origin and custom FPGA output TTL signals.

Table 3. Pupil tracking timing and latency measurements.

Camera	ROI (pix)	Exposure (ms)	Readout (ms)	Calculation FPGA (ms)	Calculation CPU (ms)	Latency (ms)
Basler	460×660	1.7	1.7	0.08	0.3	2.1
	210×300	0.7	0.8	0.04	0.3	1.2
Mikrotron	210×284	0.18	0.18	0.04	0.3	0.6

Latency is critical for stimulus delivery and eye movement stabilization because its inverse is the maximum possible correction bandwidth, which in the configurations listed in Table 3 correspond to 476, 833 and 1,667 Hz, respectively. The actual latency of a stabilization loop will be larger, due to the latency of the particular device(s) used to compensate eye movement.

3.4. Human subject pupil tracking

In order to demonstrate the proposed pupil tracker, we present two datasets collected in opposite extremes of the precision and latency ranges. The subject was not screened or selected due to any particular ocular or fixation feature that would be beneficial for pupil tracking. All user-selectable parameters had their default values and were not changed, other than for the intensity thresholding value, which was adjusted based on the live display of the camera image histogram and a live view of the thresholded image.

The first dataset, shown in Fig. 9, was collected over an ~3 s period using the high resolution Basler configuration. The camera gain was set to its minimum value (zero), which resulted in almost negligible background signal, with a maximum of one gray level (i.e., all pixel values are either 0 or 1). The illumination mask shows the LED illumination profile, which appear dark in the center at the particular distance between the LED and the paper screen. The fact that the illumination mask values span over more than an order of magnitude results in the pupil images being substantially changed by the field flattening operation, in which non-pupil dark image areas become much brighter. The annotations in this image show the edges used to fit the ellipse, the ellipse itself and its center. These annotations allow visual confirmation that some of the pixel edges shifted due to the bright Purkinje image on the left of the pupil have been successful discarded, while others have not, indicating that further algorithm refinement would be beneficial. Having said that, giving the small number of edge pixels biased by the Purkinje image it appears that the fitted ellipse is not substantially affected. A potential strategy to address this small bias, would be to implement robust ellipse fitting, meaning discarding fitting outliers and repeat the fitting one or multiple times, at the cost of increased latency.

The pupil position and rotation plots in Fig. 9 are typical of a subject fixating on a stationary target, with slow drifts separated by small saccades that are easily identified as spikes in these curves. The spectra or these time sequences show that at approximately 100 Hz, the frequency amplitudes appear to reach a noise/precision floor.

The summary of the second dataset, collected over a 5 s period using the Mikrotron camera "high frame rate" setup can be seen in Fig. 10. In this configuration, and because the background image is almost comparable to the raw image, the background subtraction has a dramatic impact. The illumination mask and the field flattening here are similar to those seen in Fig. 9. The annotations in the sample image also show the successful discarding of pixel edges due to the bright Purkinje images, but this time without clearly biased edges in the vicinity. A magnified inset of the pupil position shows the pupil position during a saccade.

Biomedical Optics EXPRESS

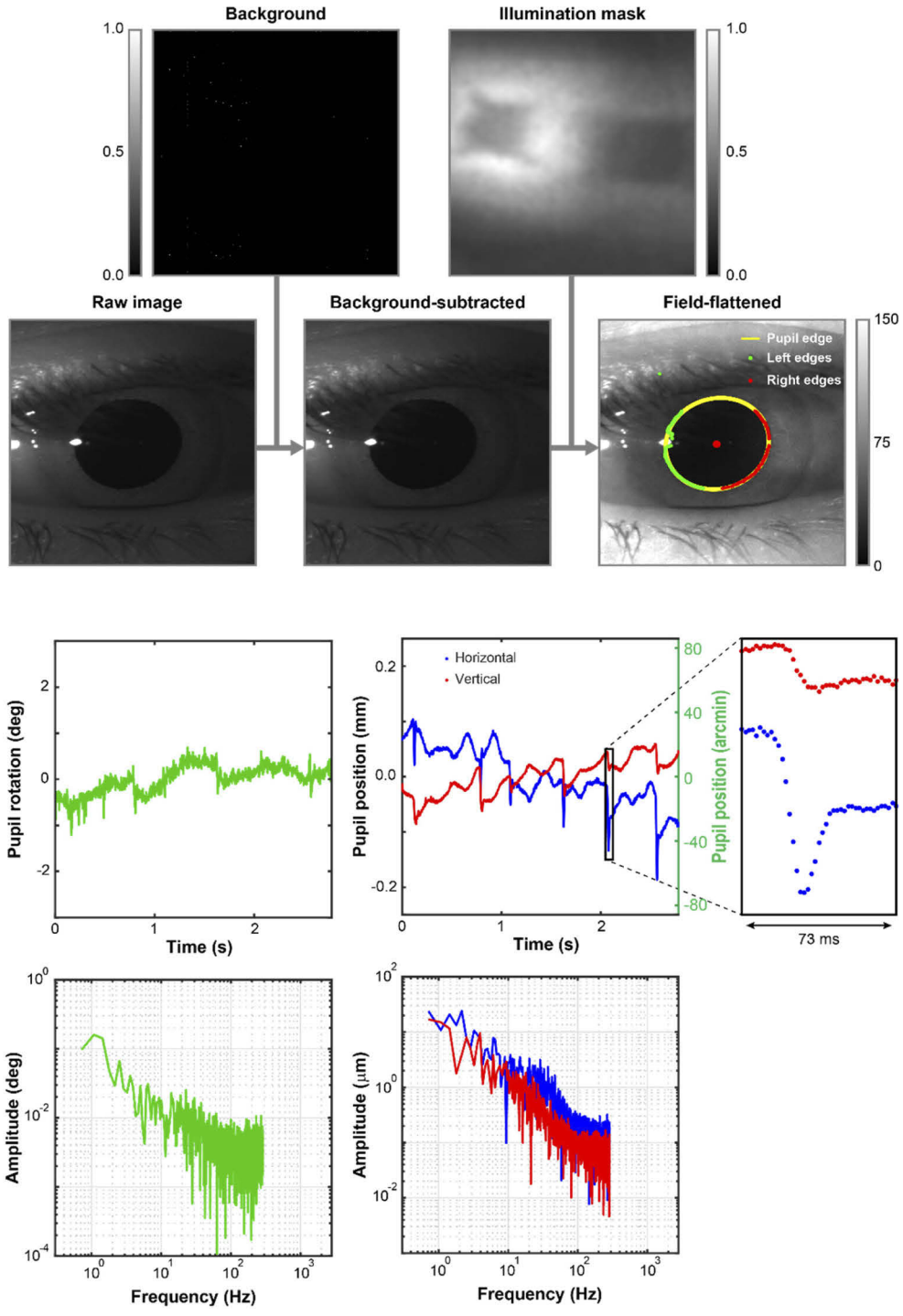

Fig. 9. Sample pupil images (18 mm field of view) and eye movement captured using the high resolution Basler camera configuration in a normal fixating subject at 575 Hz and zero camera gain. The images show the raw camera image before and after background subtraction, as well as before and after field flattening. The plots show the pupil translation and rotation across an approximately 3 s period, and their corresponding spectra.

Biomedical Optics EXPRESS

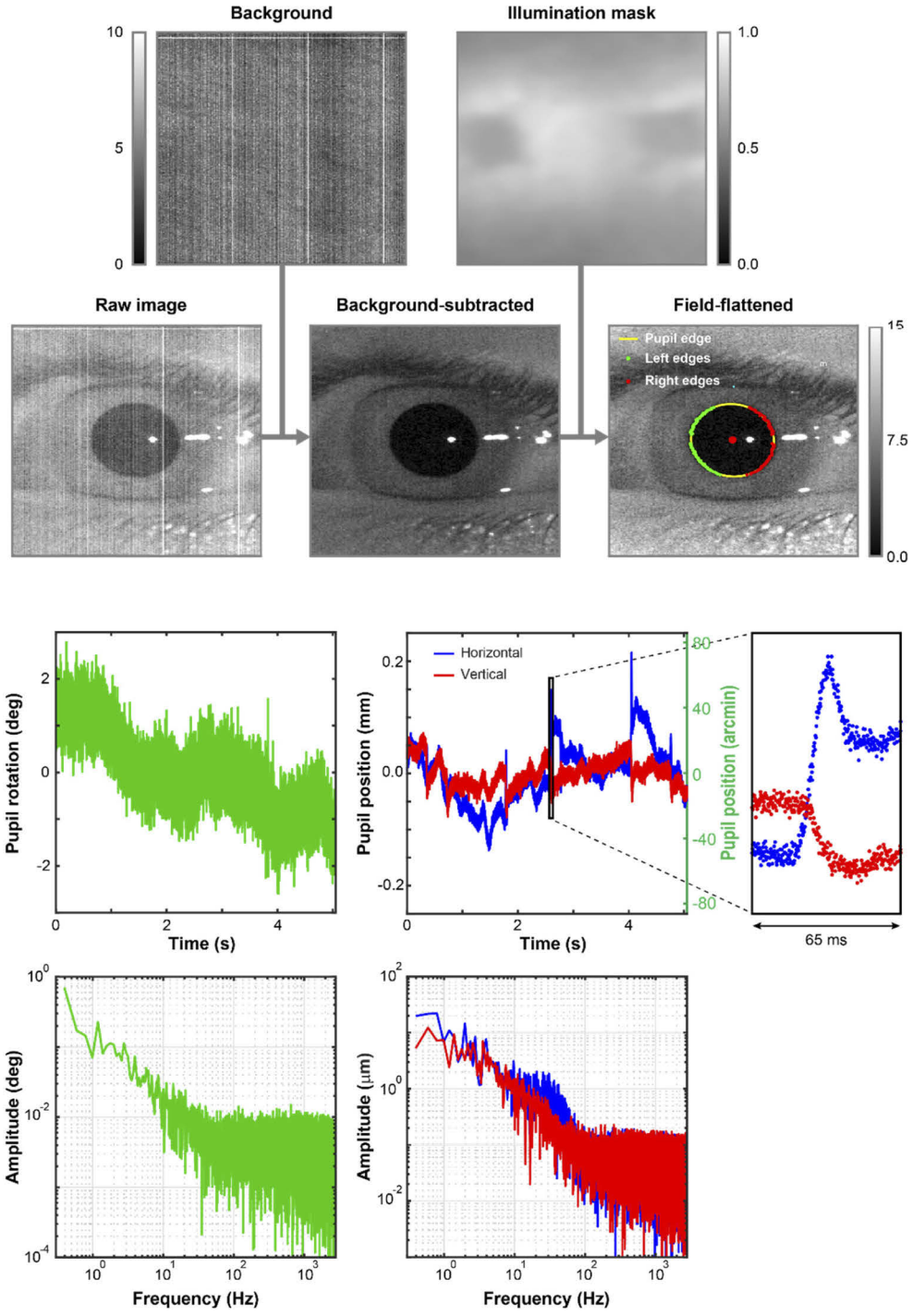

Fig. 10. Sample pupil images (18 mm field of view) and eye movement captured using the high resolution Mikrotron camera configuration in a normal fixating subject at 5400 Hz. The images show the raw camera image before and after background subtraction, as well as before and after field flattening. The plots show the pupil translation and rotation across an approximately 5 s period, and their corresponding spectra.

Biomedical Optics EXPRESS

The pupil position and rotation plots in Fig. 10 appear "thicker" than those in Fig. 9, which is indicative of larger variability due to a lower spatial sampling, an order of magnitude higher temporal sampling and lower camera sensitivity. This is quite noticeable in the magnified inset showing a saccade, with large sample-to-sample variability before, during and after the saccade. As with the previous figure, the spectra show a noise floor reached at approximately 100 Hz.

4. Summary

A low-latency monocular pupil tracker using a hybrid FPGA-CPU computing approach was described and demonstrated. This approach reduces latency by overlapping the image processing in a pixel stream with its download from the camera, as opposed to the more conventional approach in which images are fully downloaded before processing starts. The image processing consists of calculations that only require access to the values of adjacent pixels along the same image line, including background subtraction, field-flattening, thresholding, pupil edge detection and discarding of outliers, all performed on the FPGA. The final step, ellipse fitting is performed on the CPU. To illustrate that the approach is camera-agnostic, two cameras from different manufacturers were evaluated using three different optical setups. showing latencies in the 0.6-2.1 ms range with sub-pixel precision in a model pupil. Two simple tests suggest that the proposed approach works even when using only a small fraction (one eighth) of the camera 8-bit dynamic range. Pupil tracking precision in a model pupil can be as good as sub-micron and as poor as ten microns depending on the camera, optics magnification and LED power levels. These values should not be assumed to apply to pupil tracking in human subjects when object reflectivity and contrast are lower than that of the model pupil. Finally, pupil tracking was successfully demonstrated in a normal fixating subject at 575, 1250 and 5400 frames per second (1250 data not shown for brevity).

In summary, the proposed FPGA-CPU approach seems well suited for tracking the pupil with precision comparable and/or better than that of current pupil trackers and with comparable or lower latency. The high precision and low latency achieved make the device suitable for applications that require real-time eye movement compensation such as retinal imaging, retinal functional testing, retinal laser treatment and refractive surgery. Neither precision nor latency are currently limited by the FPGA or the CPU, but rather the camera quantum efficiency, SNR and download time. Should new cameras with superior specifications become available, pupil tracking precision and latency would be immediately improved.

Funding. Research to Prevent Blindness (Departmental award); National Eye Institute (P30EY026877, R01EY025231, R01EY028287, R01EY031360, R01EY032669, U01EY025477).

Acknowledgments. We would like to thank Dr. Vyas Akondi for assistance with the manuscript preparation.

Disclosures. The authors declare no conflicts of interest.

Data availability. The raw data for this work may be obtained from the authors upon reasonable request with prior approval from Stanford's institutional review board (IRB).

References

1. L. A. Riggs, J. C. Armington, and F. Ratliff, "Motions of the retinal image during fixation," J. Opt. Soc. Am. **44**(4), 315–321 (1954).
2. W. H. Hart Jr, *Adler's Physiology of the Eye: Clinical Application*, 9th ed. (Mosby-Year Book Inc., 1992).
3. N. Charman, "Optics of the Eye," in *Vision and Vision Optics*, 3rd ed., M. Bass, ed. (McGraw Hill, 2009).
4. L. F. Dell'osso and R. B. Daroff, "Congenital nystagmus waveforms and foveation strategy," Doc. Ophthalmol. **39**(1), 155–182 (1975).
5. F. Ratliff and L. A. Riggs, "Involuntary motions of the eye during monocular fixation," J. Exp. Psychol. **40**(6), 687–701 (1950).
6. M. Michaelides, J. Rha, E. W. Dees, R. C. Baraas, M. L. Wagner-Schuman, J. D. Mollon, A. M. Dubis, M. K. Andersen, T. Rosenberg, M. Larsen, A. T. Moore, and J. Carroll, "Integrity of the cone photoreceptor mosaic in oligocone trichromacy," Invest. Ophthalmol. Visual Sci. **52**(7), 4757–4764 (2011).

Biomedical Optics EXPRESS

7. M. A. Wilk, B. Higgins, R. F. Cooper, D. H. Scoles, K. E. Stepien, C. G. Summers, A. Dubra, D. M. Costakos, and J. Carroll, "Contrasting foveal specialization in disorders associated with foveal hypoplasia," Invest. Ophthalmol. Visual Sci. **55**(13), 694 (2014).
8. C. S. Langlo, E. J. Patterson, B. P. Higgins, P. Summerfelt, M. M. Razeen, L. R. Erker, M. Parker, F. T. Collison, G. A. Fishman, C. N. Kay, J. Zhang, R. G. Weleber, P. Yang, D. J. Wilson, M. E. Pennesi, B. L. Lam, J. Chiang, J. D. Chulay, A. Dubra, W. W. Hauswirth, and J. Carroll, and ACHM-001 Study Group, "Residual foveal cone structure in *CNGB3*-associated achromatopsia," Invest. Ophthalmol. Visual Sci. **57**(10), 3984–3995 (2016).
9. C. S. Langlo, L. R. Erker, M. Parker, E. J. Patterson, B. P. Higgins, P. Summerfelt, M. M. Razeen, F. T. Collison, G. A. Fishman, C. N. Kay, J. Zhang, R. G. Weleber, P. Yang, M. E. Pennesi, B. L. Lam, J. D. Chulay, A. Dubra, W. W. Hauswirth, D. J. Wilson, and J. Carroll, "Repeatability and longitudinal assessment of foveal cone structure in CNGB3-associated achromatopsia," Retina **37**(10), 1956–1966 (2017).
10. J. Lu, B. Gu, X. Wang, and Y. Zhang, "High-speed adaptive optics line scan confocal retinal imaging for human eye," PLoS One **12**(3), e0169358 (2017).
11. S. Stevenson and A. Roorda, "Correcting for miniature eye movements in high-resolution scanning laser ophthalmoscopy," Proc. SPIE **5688**, 12 (2005).
12. A. Dubra and Z. Harvey, "Registration of 2D images from fast scanning ophthalmic instruments," in *Biomedical Image Registration*, 62041 ed., B. Fischer, B. Dawant, and C. Lorenz, eds. (Springer-Verlag, Berlin, 2010), pp. 60–71.
13. D. X. Hammer, R. D. Ferguson, C. E. Bigelow, N. V. Iftimia, T. E. Ustun, and S. A. Burns, "Adaptive optics scanning laser ophthalmoscope for stabilized retinal imaging," Opt. Express **14**(8), 3354–3367 (2006).
14. Q. Yang, J. Zhang, K. Nozato, K. Saito, D. R. Williams, A. Roorda, and E. A. Rossi, "Closed-loop optical stabilization and digital image registration in adaptive optics scanning light ophthalmoscopy," Biomed. Opt. Express **5**(9), 3174–3191 (2014).
15. D. Hillmann, H. Spahr, C. Pfaffle, H. Sudkamp, G. Franke, and G. Huttmann, "In vivo optical imaging of physiological responses to photostimulation in human photoreceptors," Proc. Natl. Acad. Sci. U. S. A. **113**(46), 13138–13143 (2016).
16. R. F. Cooper, W. S. Tuten, A. Dubra, D. H. Brainard, and J. I. W. Morgan, "Non-invasive assessment of human cone photoreceptor function," Biomed. Opt. Express **8**(11), 5098–5112 (2017).
17. S. K. Cheong, W. Xiong, J. M. Strazzeri, C. L. Cepko, D. R. Williams, and W. H. Merigan, "In vivo functional imaging of retinal neurons using red and green fluorescent calcium indicators," in *Retinal Degenerative Diseases, Advances in Experimental Medicine and Biology* 1074 (Springer International Publishing, 2018), 135–144.
18. A. Joseph, A. Guevara-Torres, and J. Schallek, "Imaging single-cell blood flow in the smallest to largest vessels in the living retina," eLife **8**, e45077 (2019).
19. B. Braaf, K. V. Vienola, C. K. Sheehy, Q. Yang, K. A. Vermeer, P. Tiruveedhula, D. W. Arathorn, A. Roorda, and J. F. de Boer, "Real-time eye motion correction in phase-resolved OCT angiography with tracking SLO," Biomed. Opt. Express **4**(1), 51–65 (2013).
20. X. Hu and Q. Yang, "Modeling and optimization of closed-loop retinal motion tracking in scanning light ophthalmoscopy," J. Opt. Soc. Am. A **36**(5), 716–721 (2019).
21. C. K. Sheehy, Q. Yang, D. W. Arathorn, P. Tiruveedhula, J. F. de Boer, and A. Roorda, "High-speed, image-based eye tracking with a scanning laser ophthalmoscope," Biomed. Opt. Express **3**(10), 2611–2622 (2012).
22. C. K. Sheehy, P. Tiruveedhula, R. Sabesan, and A. Roorda, "Active eye-tracking for an adaptive optics scanning laser ophthalmoscope," Biomed. Opt. Express **6**(7), 2412–2423 (2015).
23. H. D. Crane and C. M. Steele, "Accurate three-dimensional eyetracker," Appl. Opt. **17**(5), 691–714 (1978).
24. H. D. Crane and C. M. Steele, "Generation-V dual-Purkinje-image eyetracker," Appl. Opt. **24**(4), 527–537 (1985).
25. A. T. Duchowski, "A breadth-first survey of eye-tracking applications," Behav. Res. Methods **34**(4), 455–470 (2002).
26. C. H. Morimoto and M. R. M. Mimica, "Eye gaze tracking techniques for interactive applications," Comput. Vis. Image Underst **98**(1), 4–24 (2005).
27. M. Wedel, "A review of eye-tracking research in marketing," in *Review of Marketing Research*, 4R. Pieters and K. M. Naresh, eds. (Emerald Group Publishing Limited, 2008), pp. 123–147.
28. T. T. Brunyé, T. Drew, D. L. Weaver, and J. G. Elmore, "A review of eye tracking for understanding and improving diagnostic interpretation," Cogn. Research **4**(1), 7 (2019).
29. A. H. Clarke, J. Ditterich, K. Drüen, U. Schönfeld, and C. Steineke, "Using high frame rate CMOS sensors for three-dimensional eye tracking," Behav. Res. Methods **34**(4), 549–560 (2002).
30. O. Carrasco-Zevallos, D. Nankivil, B. Keller, C. Viehland, B. J. Lujan, and J. A. Izatt, "Pupil tracking optical coherence tomography for precise control of pupil entry position," Biomed. Opt. Express **6**(9), 3405–3419 (2015).
31. O. M. Carrasco-Zevallos, D. Nankivil, C. Viehland, B. Keller, and J. A. Izatt, "Pupil tracking for real-time motion corrected anterior segment optical coherence tomography," PLoS One **11**(8), e0162015 (2016).
32. X. Wang, M. R. Beotra, T. A. Tun, M. Baskaran, S. Perera, T. Aung, N. G. Strouthidis, D. Milea, and M. J. A. Girard, "In Vivo 3-dimensional strain mapping confirms large optic nerve head deformations following horizontal eye movements," Invest. Ophthalmol. Visual Sci. **57**(13), 5825–5833 (2016).
33. A. N. Kuo, P. K. Verkicharla, R. P. McNabb, C. Y. Cheung, S. Hilal, S. Farsiu, C. Chen, T. Y. Wong, M. K. Ikram, C. Y. Cheng, T. L. Young, S. M. Saw, and J. A. Izatt, "Posterior eye shape measurement with retinal OCT compared to MRI," Invest. Ophthalmol. Visual Sci. **57**(9), OCT196 (2016).

Biomedical Optics EXPRESS

34. H. Deubel and B. Bridgeman, "Fourth Purkinje image signals reveal eye-lens deviations and retinal image distortions during saccades," Vision Res. **35**(4), 529–538 (1995).
35. J. Tabernero and P. Artal, "Lens oscillations in the human eye. Implications for post-saccadic suppression of vision," PLoS One **9**(4), e95764 (2014).
36. J. J. Vos, "Colormetric and photometric properties of a 2 degree fundamental observer," Color Res. Appl. **3**(3), 125–128 (1978).
37. ANSI, "American national standard for safe use of lasers ANSI Z136.1 - 2014," (2014).
38. Y. Yang, K. Thompson, and S. A. Burns, "Pupil location under mesopic, photopic, and pharmacologically dilated conditions," Invest. Ophthalmol. Visual Sci. **43**(7), 2508–2512 (2002).
39. T. Y. Wong, P. J. Foster, T. P. Ng, J. M. Tielsch, G. J. Johnson, and S. K. L. Seah, "Variations in ocular biometry in an adult chinese population in Singapore: The Tanjong Pagar Survey," Invest. Ophthalmol. Visual Sci. **42**(1), 73–80 (2001).
40. E. Ojaimi, K. A. Rose, I. G. Morgan, W. Smith, F. J. Martin, A. Kifley, D. Robaei, and P. Mitchell, "Distribution of ocular biometric parameters and refraction in a population-based study of Australian children," Invest. Ophthalmol. Visual Sci. **46**(8), 2748–2754 (2005).
41. R. Fotedar, J. J. Wang, G. Burlutsky, I. G. Morgan, K. Rose, T. Y. Wong, and P. MIthcell, "Distribution of axial length and ocular biometry measured using partial coherence laser interferometry (IOL Master) in an older white population," Ophthalmology **117**(3), 417–423 (2010).
42. H. Hashemi, M. Khabazkhoob, M. Miraftab, M. H. Emamian, M. Shariati, T. Abdolahinia, and A. Fotouhi, "The distribution of axial length, anterior chamber depth, lens thickness, and vitreous chamber depth in an adult population of Shahroud, Iran," BMC Ophthalmol. **12**(1), 50 (2012).
43. "OpenGL", retrieved https://www.opengl.org/.
44. E. Arias-Castro and D. L. Donoho, "Does median filtering truly preserve edges better than linear filtering?" Ann. Statist. **37**(3), 1172–1206 (2009).
45. J. Kolodko, S. Suzuki, and F. Harashima, "Eye-gaze tracking: an approach to pupil tracking targeted to FPGAs," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems* (2005), 344–349.
46. S. I. Kim, J. M. Cho, J. Y. Jung, S. H. Kim, J. H. Lim, T. W. Nam, and J. H. Kim, "A fast center of pupil detection algorithm for VOG-based eye movement tracking," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference* (2005), 3188–3191.
47. T. Ando, V. G. Moshnyaga, and K. Hashimoto, "A low-power FPGA implementation of eye tracking," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP, 2012), 1573–1576.
48. D. A. Padilla, J. A. B. Adriano, J. R. Balbin, I. G. Matala, J. J. R. Nicolas, and S. R. R. Villadelgado, "Implementation of eye gaze tracking technique on FPGA-based on-screen keyboard system using verilog and MATLAB," in *TENCON 2017 - 2017 IEEE Region 10 Conference* (2017), 2771–2776.
49. K. Dohi, Y. Hatanaka, K. Negi, Y. Shibata, and K. Oguri, "Deep-pipelined FPGA implementation of ellipse estimation for eye tracking," in *22nd International Conference on Field Programmable Logic and Applications (FPL)* (2012), 458–463.
50. H. Qin, X. Xu, Z. Hu, and D. Zhang, "Eye tracking system based on SOPC," in *2014 21st IEEE International Conference on Electronics, Circuits and Systems* (ICECS), 2014), 171–174.
51. L. N. Thibos, R. A. Applegate, J. T. Schwiegerling, and R. Webb, "Standards for reporting the optical aberrations of eyes," J. Refract. Surg. **18**(5), S652–S660 (2002).
52. E. Anderson, Z. Bai, C. Bischof, L. S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, and A. McKenney, *LAPACK users' guide* (SIAM, 1999).
53. "Ellipse", retrieved https://en.wikipedia.org/wiki/Ellipse.